Claims

What is claimed is:

Independent claim 1

1.   In a system comprising a plurality of perspectives and a plurality of categories, wherein at least one category is associated with a perspective to reflect associations among related categories, a method for simultaneously classifying at least one document into a plurality of categories, said method comprising:

associating a plurality of category features with each said category, wherein each of said category features represents one of a plurality of tokens;

producing a category vector for each of said plurality of categories, wherein each category vector includes said plurality of category features with a weight corresponding to each category feature, said weight indicative of a degree of association between said category feature and said category;

associating a plurality of document features with each said document, wherein each of said document features represents one of a plurality of tokens found in said document;

producing a feature vector for each said document, wherein each feature vector includes said plurality of document features with a count corresponding to each document feature, said count indicative of the number of times said document feature appears in said document;

multiplying said category vector by said document vector, in accordance with the mathematical convention of multiplication of a vector by a vector, to produce a plurality of category scores for each document; and

for each perspective, classifying a document into a category provided said category score exceeds a predetermined threshold.

2.   The method of claim 1, wherein the producing a feature vector step further comprises producing a feature vector for each said document, wherein each feature vector includes said plurality of document features with a count corresponding to each document feature, said count indicative of the number of times said document feature appears in said document, provided said document feature is not a stop word.

3.   The method of claim 1, wherein the producing a feature vector step further comprises producing a feature vector for each said document, wherein each feature vector includes said plurality of document features with a count corresponding to each document feature, said count indicative of the number of times said document feature appears in said document, provided said document feature is not FilteredOut.

Independent claim 1a

4.   The method of claim 1, wherein the classifying step further comprises:

comparing a category score for a first document in a first perspective with a category score for said first document in a second perspective; and

modifying the category score in the first perspective in response to the category score in the second perspective.

Independent claim 1.a.1

5.   The method of claim 4, wherein the modifying step further comprises:

determining whether a category score for a first document in a first perspective exceeds a predetermined threshold for said first perspective; and

flagging a second document as a DirectHit document for a category in a second perspective, provided said first document exceeds said predetermined threshold.

6.   The method of claim 4, wherein said first perspective is an ancestor of said second perspective.

7.   The method of claim 4, wherein the modifying step further comprises:

determining whether a category score for a first document in a first perspective exceeds a predetermined threshold for said first perspective; and

flagging a second document as a RejectConcept document for a category in a second perspective, provided said first document does not exceed said predetermined threshold.

Independent claim 1.a.3

8.   The method of claim 4, wherein the comparing step further comprises creating an ordered list of category scores for each category in each perspective;

9.   The method of claim 8, wherein the comparing step further comprises identifying a document associated with a highest category score in a first perspective;

10.   The method of claim 9, wherein the modifying step further comprises:

locating the document in a second perspective; and

decreasing the category score associated with said document in said second perspective, provided said category score is not a highest category score in said second perspective.

11.   The method of claim 8, further comprising the step of repeating the locating and decreasing steps for every perspective.

12.   The method of claim 1, wherein the weight corresponding to said concept node feature is between -1 and 1.

13.   A method for associating at least one of a plurality of features with at least one of a plurality of categories, said method comprising at least one of manually or automatically associating at least one of said plurality of features to at least a first category, said plurality of features contributing to a decision to classify a document into said at least first category.

14.   The method of claim 13, further including classifying at least one document into said at least one category, provided the document includes a predetermined number of said plurality of features associated with said category.

15.   The method of claim 13, further comprising at least one of manually or automatically associating at least one of a plurality of attributes with at least one of said plurality of features, said plurality of attributes contributing to a decision to classify a document into said at least one category.

16.   The method of claim 15, further comprising:

determining whether said at least one feature was manually associated to said at least first category; and

associating an attribute with said at least one feature that indicates that the feature was Edited.

17.   The method of claim 16, further comprising at least one of manually or automatically associating at least one of said plurality of features from said first category to a second category, provided the feature does not contain an attribute associated with said at least first category that declares the feature to be Edited.

18.   The method of claim 17, further comprising manually associating at least one of said plurality of features from said first category to a second category, provided the feature contains an attribute associated with said at least first category that declares the feature to be Edited.

19.   The method of claim 15, further comprising classifying a document into a category, provided the document does not contain a feature whose association with said category has a RejectConcept attribute.

20.   The method of claim 15, further comprising classifying a document that contains a feature or a morphological variant of the feature into a category, provided the feature contains an attribute associated with the category that declares the feature to be morphologically variable.

21.   The method of claim 15, further comprising classifying a document that contains a feature into a category, provided the feature contains an attribute associated with the category that declares the feature to be morphologically invariant.

22.   The method of claim 15, further comprising the step of not classifying a document that contains a morphological variant of a feature into a category, provided the feature contains an attribute associated with the category that declares the feature to be morphologically invariant.

23.   The method of claim 15, further comprising classifying a document that contains a feature or a case variant of the feature into a category, provided the feature contains an attribute associated with the category that declares the feature to be case insensitive. a feature associated with a category which association has an attribute declaring the feature to be case insensitive contributes to a decision to classify a document into said category, provided the document contains the feature or a case variant of the feature.

24.   The method of claim 15, further comprising classifying a document that contains a feature into a category, provided the feature contains an attribute associated with the category that declares the feature to be case insensitive.

25.   The method of claim 15, further comprising the step of not classifying a document that contains a case variant of a feature into a category,

provided the feature contains an attribute associated with the category that declares the feature to be case invariant

26. The method of claim 15, further comprising classifying at least one document into at least one of said categories, provided the document contains a feature whose association with said at least one category has an attribute entitled DirectHit.

27. The method of claim 15, further comprising classifying at least one document into at least one of said categories, provided the document does not contain a feature whose association with said at least one category does not contain the attribute entitled Overlap.

28. The method of claim 15, further comprising classifying a document containing an overlapping feature into a category, provided the feature contains an attribute associated with the category that declares the feature to be overlap insensitive.

29. The method of claim 15, further comprising classifying a document containing a non-overlapping feature into a category, provided the feature contains an attribute associated with the category that declares the feature to be overlap sensitive.

30. The method of claim 15, further comprising the step of not classifying a document that contains an overlapping feature into a category, provided the feature contains an attribute associated with the category that declares the feature to be overlap sensitive.

31. The method of claim 15, further comprising at least one of manually or automatically assigning a weight to the feature, said weight indicative of a degree of association between said document and said category.

32. The method of claim 31, further comprising:

determining whether said weight was manually assigned to said feature; and

associating an attribute with said feature that indicates that the weight was WeightEdited.

33.   The method of claim 32, further comprising at least one of manually or automatically replacing a value for said weight with another value, provided the feature does not contains an attribute associated with the category that declares the feature to be WeightEdited.

34.   The method of claim 33, further comprising manually replacing a value for said weight with another value, provided the feature contains an attribute associated with the category that declares the feature to be WeightEdited.

35.   The method of claim 15, further comprising classifying a document containing an overlapping feature into a category, provided the feature contains an attribute associated with the category that declares the feature to be overlap insensitive.

36.   The method of claim 15, further comprising classifying a document containing a non-overlapping feature into a category, provided the feature contains an attribute associated with the category that declares the feature to be overlap sensitive.

37.   The method of claim 15, further comprising the step of not classifying a document that contains an overlapping feature into a category, provided the feature contains an attribute associated with the category that declares the feature to be overlap sensitive.

38.   The method of claim 15, further comprising the steps of:

determining whether at least one of said plurality of features is a stop word; and

setting an attribute indicating that said feature is a stop word.

39.   The method of claim 15, further comprising the steps of:

at least one of manually or automatically determining a scope of at least one of said plurality of features; and

setting an attribute indicating that said at least one feature is for queries only, or for documents only, or for both.

40.   The method of claim 15, further comprising the step of setting an attribute indicating that said feature is FilteredOut, provided said feature has been manually or automatically filtered out of a classification.

41.   The method of claim 31, further comprising multiplying said weight by a scaling parameter, provided the decision to classify the document into said category was based on at least one feature automatically associated with the category.

42.   The method of claim 41, wherein said scaling parameter is between 0 and 1.

43.   A method for associating one of a plurality of features with at least one of a plurality of perspectives, wherein at least one feature is associated with at least one category, and at least one of said categories is associated with at least one of said plurality of perspectives to reflect associations among related categories and features, said method comprising at least one of manually or automatically associating a plurality of features to at least one perspective, said plurality of features contributing to a decision to classify a document into said at least one category.

44.   The method of claim 13, further comprising the steps of:

reviewing at least one automatic association of a feature to a category; and

characterizing said association as a manual association;

wherein a manual association contributes greater to a decision to classify a feature into said category.